Population-based Approaches to Gene Discovery

Lisa Cannon Albright, PhD
Professor, Chief
Division of Genetic Epidemiology
Department of Internal Medicine
University of Utah School of Medicine

Today’s Outline

- The Utah Genealogy resource (UPDB)
- Heritable Contribution to Health-Related Traits
- Utah Gene Discovery Approaches
- The Utah Future

Today’ s Outline

- **The Utah Genealogy resource (UPDB)**
- **Heritable Contribution to Health-Related Traits**
- **Utah Gene Discovery Approaches**
- **The Utah Future**

Utah Genealogy - UPDB

In the 1960’s Mark Skolnick, PhD constructed a computerized genealogy in the Parma Valley from parish records; medical records were diffuse in this population

In the 1970’s he proposed building the Utah Genealogy - Utah genealogy and linked phenotype data

In the 1990s he co-founded Myriad Genetics

The Utah Population

Adherence to proscriptions against coffee, tea, tobacco, alcohol high

Teachings encourage large families with strong ties, high educational attainment, and strict sexual mores

During the late nineteenth century, mean number of offspring per couple was >5 during each decade and reached almost 9

Before 1890 polygamy was practiced among the pioneers, leading to families with multiple wives and dozens of children

Founding pioneers were largely unrelated; low/normal inbreeding

Receptive to research studies

Original Utah Genealogy Data

- **Mormons make up 75% of the state of Utah**
  - **individual Mormons trace their ancestries as far as possible**
  - **records collected in the Family History Library of the Church**
- **The Utah Genealogy used 3-generation family genealogy sheets submitted by members of the Church of Jesus Christ of Latter-day Saints**
- **Skolnick selected sheets containing at least one life event in Utah or on the pioneer trail (1840-1850); record linking accomplished during data entry**
- **Original Utah genealogy included 1.6 million individuals linked in genealogies 6 - 7 generations deep**
- **The genealogy is now extended with Utah vital statistics data**

Utah Genealogy and Diagnosis Resources

- **Genealogy of ~2.5 million Utah Mormon pioneers and descendants**
  - **Linked to Utah cancer registry records (1966 - present)**
  - **Linked to Utah death certificate records (1904 - present)**
  - **Other linked data**
    - Hospital Discharge data for the state
      - contains over 3 million claims
    - University of Utah Health Sciences Center (UUHSC) Enterprise Data Resource
      - contains over 1.4 million patient demographic records
    - Intermountain Healthcare EDW (largest health care provider in Utah)
      - Contains over 3.4 million patient demographic records

Example UPDB Phenotype Data Today

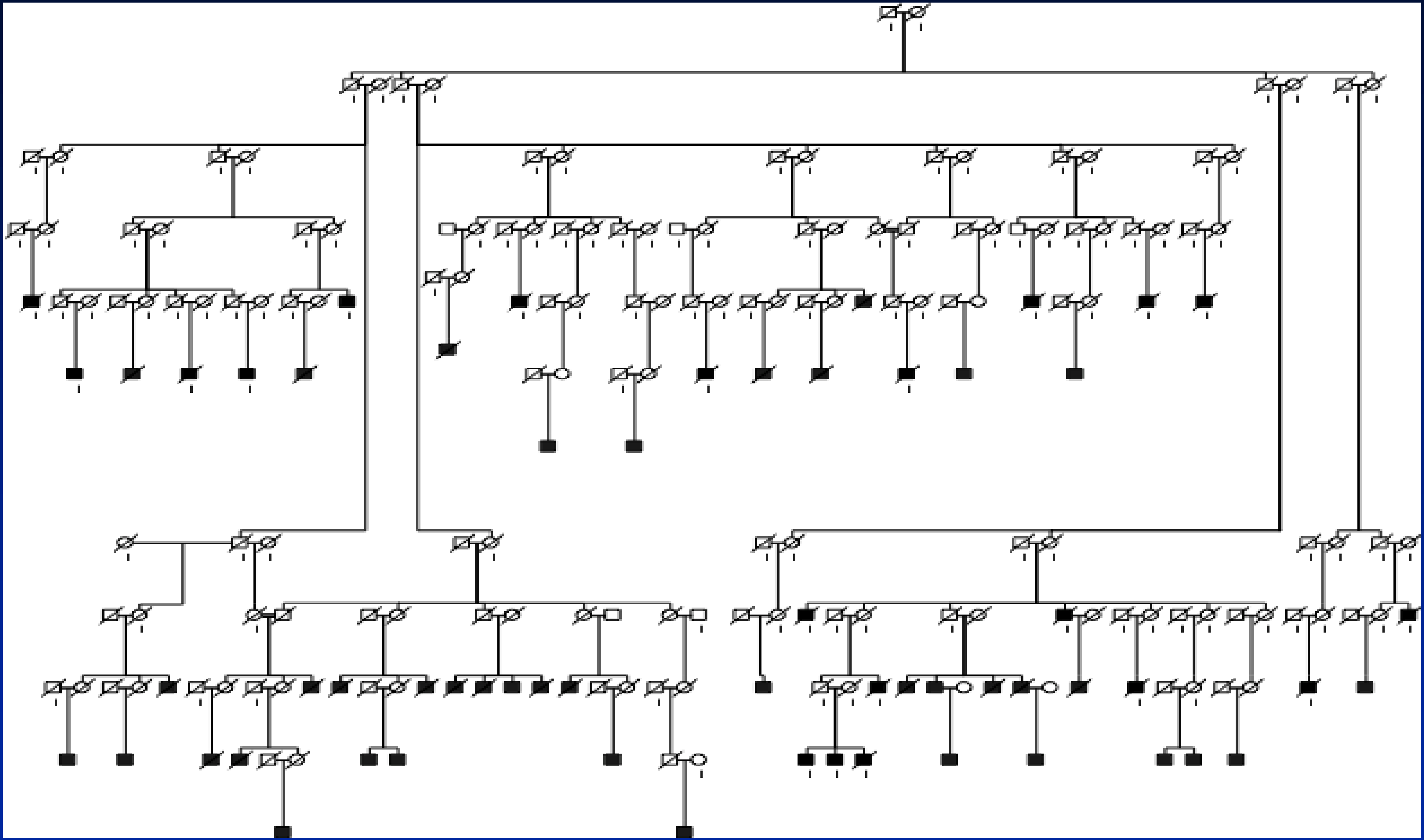
Death Certificates	**722,000: age and causes of death back to 1904**
Cancer Records	**234,000+ NCI SEER registry records from 1966**
Hospitals	**2-4 million patients: diagnosis, procedures, medications, response, from 1993**
Driver’s License	**2.8 million individuals: Body Mass Index**
Birth Certificates	**2.2 million births: birth weight, APGAR, gestational diabetes, …**
Other	**longevity, fertility**

Records available in UPDB today

Record Type	**Number of Records**
Genealogy Records	1,602,035
1880 Census	142,711
Birth Certificates (1915-21, 1947-2006)	2,139,399
Marriage Certificates (1978-2005)	569,556
Divorce Records (1978-2005)	249,189
Fetal Deaths	7,790
Social Security Death Index	479,474
Idaho Cancer Records (1969-2005)	123,737
Inpatient Hospital Claims (1996-2006)	2,726,004
…	
Total Records	**11.8 million**

pedigrees now 15 generations

Sample Utah Prostate Pedigree (60 cases)



Today's Outline

- **The Utah Genealogy resource (UPDB)**
- **Heritable Contribution to Health-Related Traits**
- **Utah Gene Discovery Approaches**
- **The Utah Future**

Identification of a Heritable Contribution to any health-related trait with UPDB data

We use 3 different methods to evaluate the genetic contribution to a phenotype using the UPDB:

1. **Familial clustering (average relatedness) of affecteds**
2. **Relative risk of disease in relatives of affecteds**
3. **Identification of pedigrees with an excess of affecteds**

Identification of a Heritable Contribution to any health-related trait

We use 3 different methods to evaluate the genetic contribution to a phenotype using the UPDB:

1. **Familial clustering (average relatedness) of affecteds**
2. **Relative risk of disease in relatives of affecteds**
3. **Identify pedigrees with an excess of affecteds**

Utah Analysis of Familial Clustering

Concept: For a disease with a familial/genetic component, cases should be more closely related than randomly selected matched controls

Method: Using a genealogy linked to diagnosis you can define the genetic relationships between all the cases, and estimate the average “relatedness” of all individuals with a specific disease.

The same measurement of relatedness on a set of matched controls estimates the average relatedness expected in the population

Utah Analysis of Familial Clustering

Interpretation: If the average “relatedness” of all of the cases is significantly higher than that of the controls, you have evidence of excess familial clustering

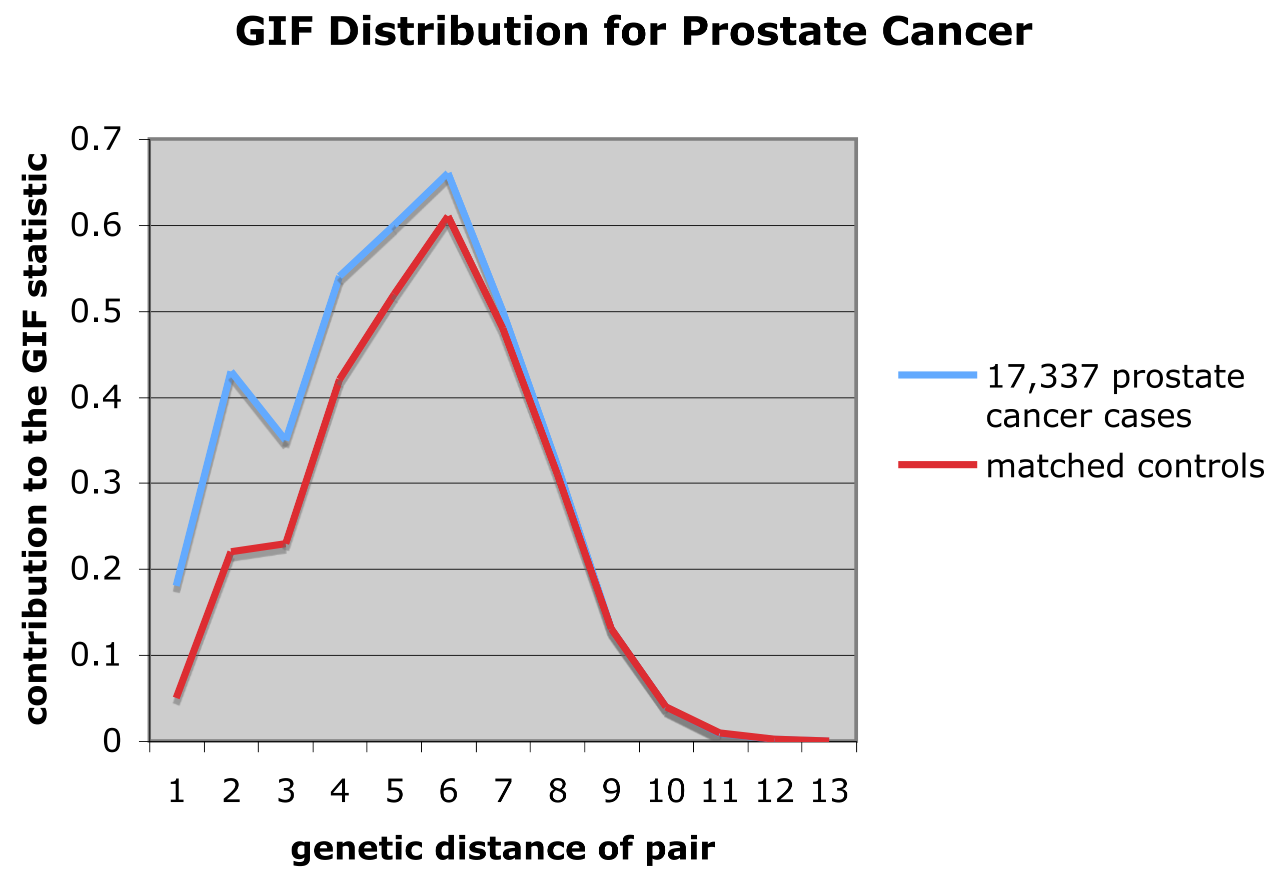
But is the excess familial clustering genetic or environmental?

If the excess of relationships is not limited to close relationships, an inherited predisposition is more likely than shared familial or environmental effects

Utah Analysis of Familial Clustering by Disease

- Stroke
- Aneurysm mortality
- Coronary Heart Disease
- Lobular breast cancer
- Prostate Cancer
- Melanoma
- Influenza mortality
- Chronic Fatigue Syndrome
- Asthma mortality
- Surgical site infection
- traumatic injury
- osteoarthritis
- Parkinsons
- ALS
- Alzheimers
- Brain cancer
- Coronary valve disease
- Benign pituitary tumors
- Rotator cuff disease
- Pancreas cancer
- Colon cancer
- Rectal cancer
- Cervical spondylotic myelopathy
- Pelvic organ prolapse
- Ataxias
- Sarcoma
- leanness
- pulmonary fibrosis
- chiari malformations
- Endometrial cancer
- Diabetes
- Lumbar disc disease
- Myeloma

Utah Analysis of Prostate Cancer Familial Clustering

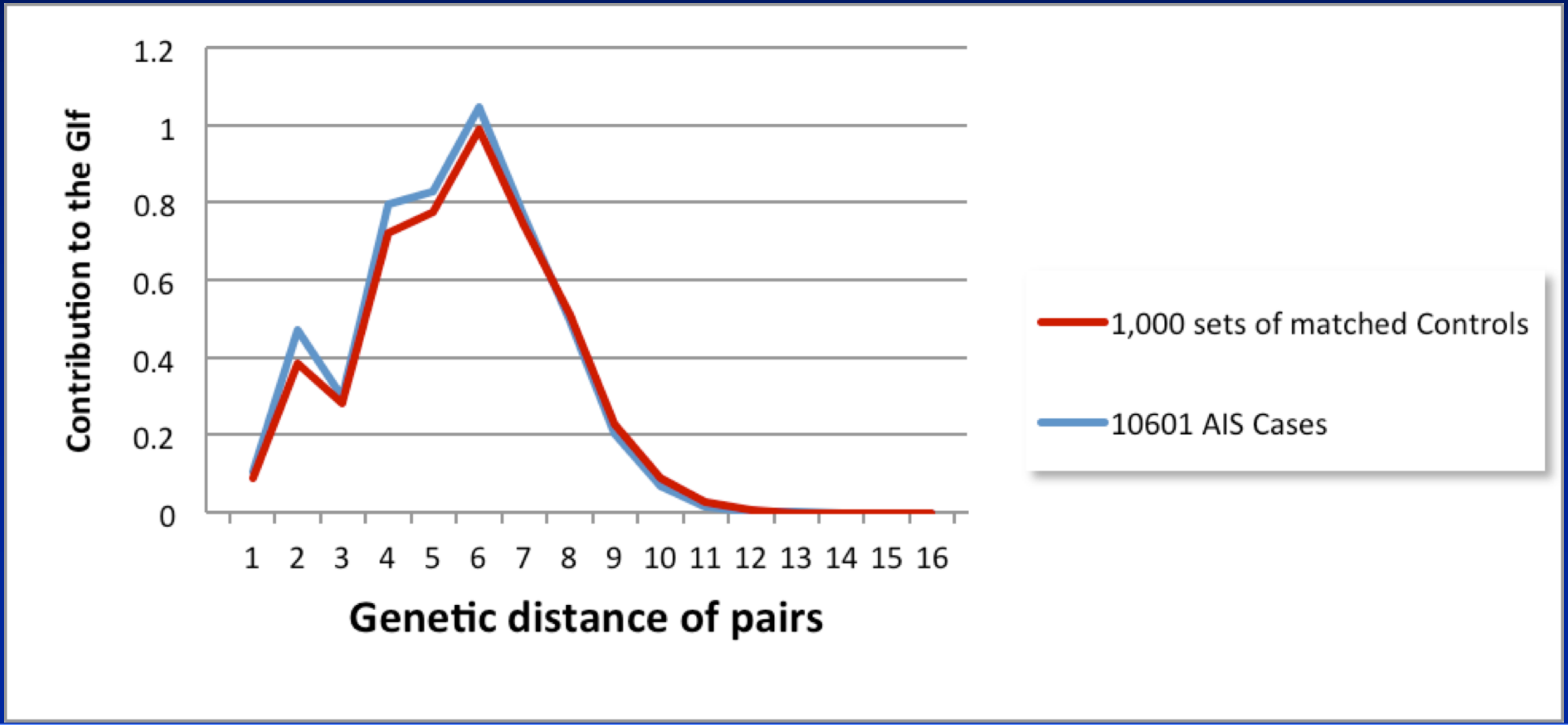


Utah Analysis of Stroke Death Clustering

Cause of death = Ischemic Stroke (before 1980)

overall excess relatedness $p<0.001$

excess relatedness in distant relationships only $p=0.032$



Identification of a Heritable Contribution to any health-related trait

We use 3 different methods to evaluate the genetic contribution to a phenotype using the UPDB:

1. **Familial Clustering (average relatedness) of affecteds**
2. **Relative risk of disease in relatives of affecteds**
3. **Identify pedigrees with an excess of affecteds**

Relative Risk in Relatives

Concept: Diseases with a heritable component should occur at higher than population frequency among relatives of cases.

Method: Compare rate of phenotype in relatives of cases to the population rate (estimated in the UPDB)

Interpretation: Excess first degree relative risks may be due to genes, environment, or both

Excess risks in more distant relatives strongly support a heritable component

First, Second, and Third-Degree Relationships

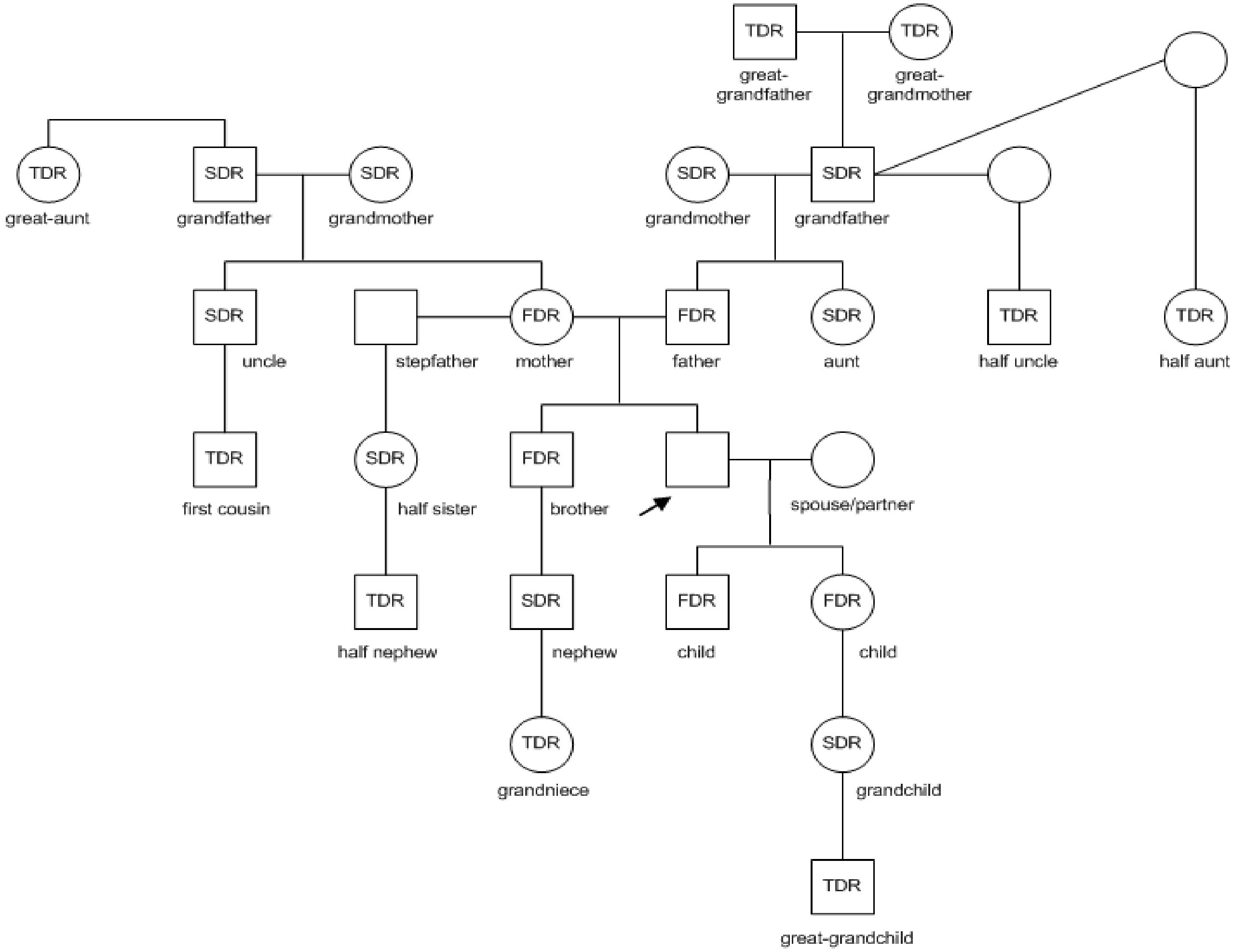


Figure 1: Pedigree diagram illustrating family relationships and degrees. The proband is indicated with an arrow. FDR = first-degree relative, SDR = second-degree relative, and TDR = third-degree relative.

First Degree Relative Risk for Intracranial Aneurysms

581 Intracranial Aneurysm (IA) deaths in the UPDB

5,280 of their first degree relatives also have a death certificate

	Obs	Exp	RR	p value
Intracranial Aneurysm	19	4.7	4.05	<0.0001

Cannon Albright et al 2003

Relative Risks for death from Stroke

death after 1980 and age at death < 75 years (youngest 20% of deaths)

1,812 deaths from ischemic stroke before 1980 and before age 75 years in the UPDB

6,458 of their first degree relatives have a death certificate
14,921 of their second degree relatives have a death certificate
36,709 of their third degree relatives have a death certificate

	Obs	Exp	RR	95% CI
1st degree relatives stroke death	298	194.5	1.53	1.36, 1.72
2nd degree relatives stroke death	486	329.1	1.48	1.35, 1.61
3rd degree relatives stroke death	1418	963.9	1.47	1.40, 1.55

(Majersik et al., submitted)

Identification of a Heritable Contribution to any health-related trait

We use 3 different methods to evaluate the genetic contribution to a phenotype using the UPDB:

1. **Familial Clustering (average relatedness) of affecteds**
2. **Relative risk of disease in relatives of affecteds**
3. **Identify pedigrees with an excess of affecteds**

Pedigrees with a significant excess of a selected phenotype

Concept: if the descendants of a couple have a higher number of cases than expected based on population rates, the pedigree is high-risk

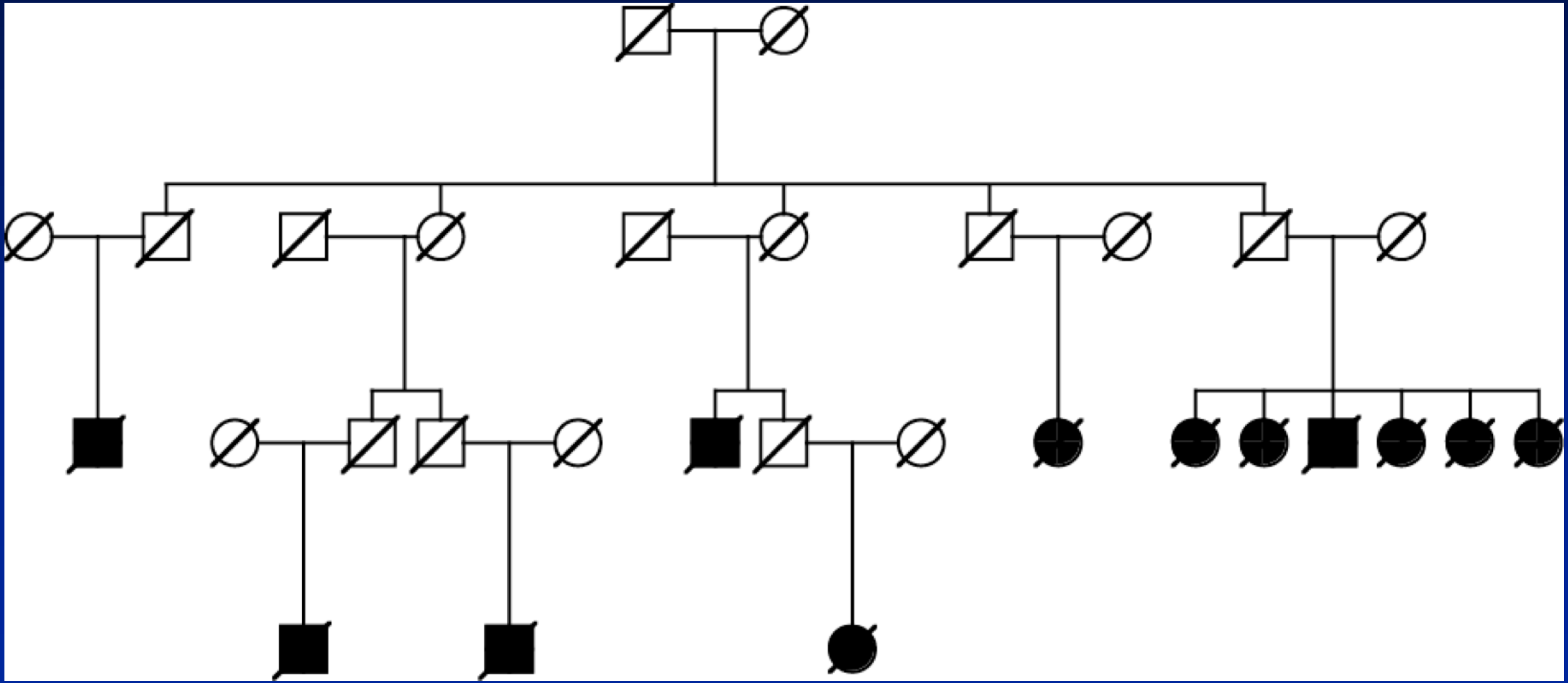
Method: Identify cases in the UPDB

Compare observed and expected numbers of cases among descendants using internally estimated rates

Significance: Typically not possible for other groups to identify "high-risk" pedigrees (as opposed to pedigrees with lots of related cases)

Utah Pedigree with an excess of death from Stroke

# descendants	phenotype	observed	expected	p-value
~2400	stroke	12	4.6	.0006



Today’s Outline

- **The Utah Genealogy resource (UPDB)**
- **Heritable Contribution to Health-Related Traits**
- **Utah Gene Discovery Approaches**
  - **Established**
  - **New**
- **The Utah Future**

Genes Localized/Isolated by the Genetic Epidemiology group using the UPDB high-risk pedigrees

1987	**Neurofibromatosis**
1988	**Alport Syndrome**
1994	*CDKN2A*- ***p16*** - melanoma
1994	***BRCA1***
1996	***BRCA2***
2001	***HPC2/ELAC2*** - prostate cancer
2014	Melanoma predisposition gene???

Utah High Risk Pedigree Studies

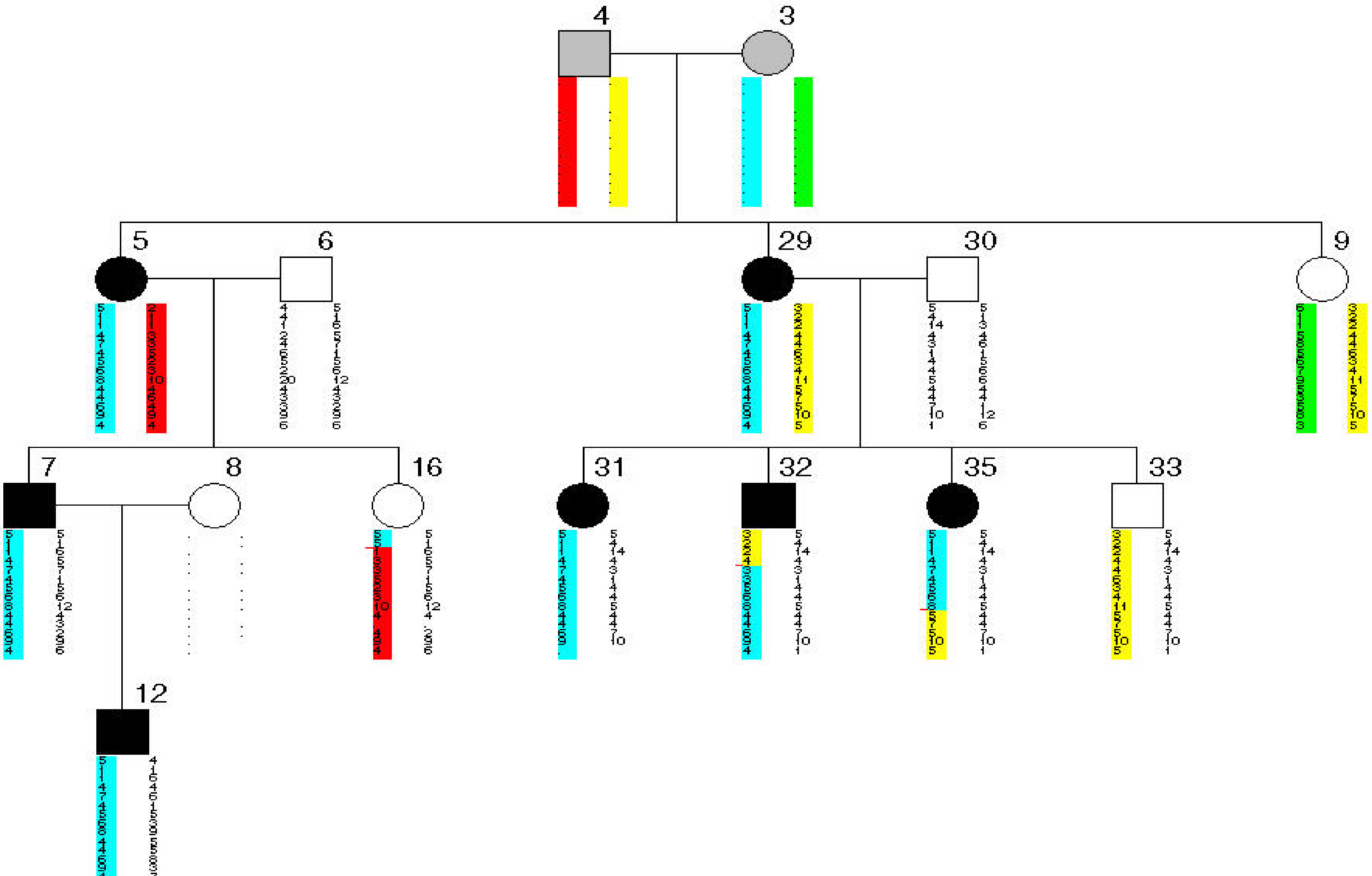
Disease	**Time period**	**Pedigrees**	**DNAs**
Breast Cancer	1972 - 1992	490	7536
Colon Cancer	1981-1991	272	4112
Celiac	1984 – present	319	2351
Melanoma	1986 - present	179	3075
Prostate Cancer	1990 - present	562	9737
Asthma	1995 - 2000	198	2494
Osteoporosis	1995 - 2000	323	1918
Depression	1996 - 2000	438	2939
Intracranial Aneurysm	1998 - present	151	721
Chronic Lymphocytic Leukemia	2008 – present	150	600
Multiple Myeloma	2008 - present	150	600
Pelvic Organ Prolapse	2010 – present	83	548
Pancreas Cancer	2010 – present	107	180

Total 36,000+ DNAs

Utah Gene Discovery Approaches - Linkage Analysis in Extended Pedigrees

- Identify high-risk pedigrees
- Sample affected individuals and informative relatives
- Genotype informative pedigree members
- Analyze co-segregation of markers and phenotype to identify identical regions of chromosomes shared by cases
- Identify pedigrees exhibiting linkage evidence (LOD > 0.588)
- Identify chromosomal region shared by cases in linked pedigrees
- Minimize region defined by *all* pedigrees

Analyzing Cosegregation of markers in Linked Pedigrees to Localize Gene

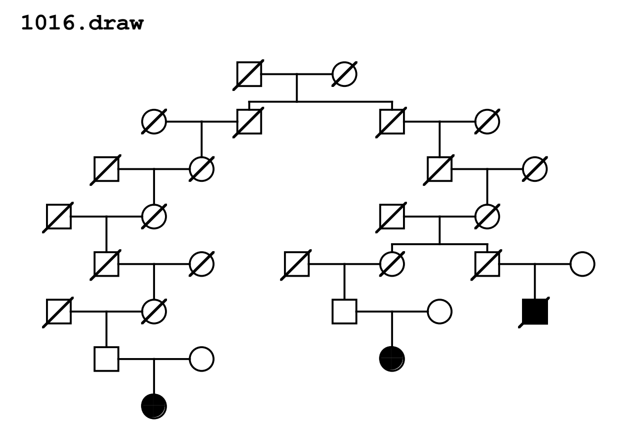


Today’s Outline

- The Utah Genealogy resource (UPDB)
- Heritable Contribution to Health-Related Traits
- Utah Gene Discovery Approaches
  - Established
  - New
- The Utah Future

New Utah Gene Discovery Approaches

Search for chromosomal sharing Identical by Descent (IBD) in distantly related cases (~ linkage)



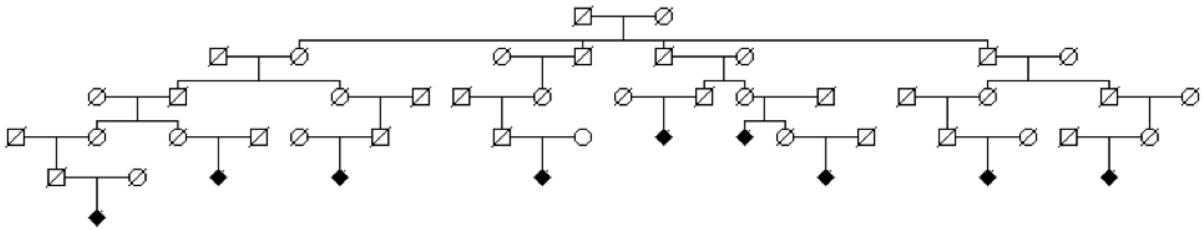
Distant relatives will share genomic regions inherited from a common ancestor, with an expected length of sharing based on distance. Sharing of longer segments than expected (and a phenotype) suggests they are cosegregating

Powerful approach for rare phenotypes

New Utah Gene Discovery Approaches

Use stored FFPE tissues to obtain germline DNA for deceased Cases—powerful for fatal diseases

The founder of this pedigree has over 4,800 descendants in the UPDB; 9 descendants have had pancreatic cancer; we expected only 2.6 cases among the descendants (p=0.0004).



Today’s Outline

- **The Utah Genealogy resource (UPDB)**
- **Heritable Contribution to Health-Related Traits**
- **Utah Gene Discovery Approaches**
- **The Utah Future**

Utah Genealogical Resource Potential Future

Expand boundaries of the Utah Genealogy
- genealogy of western states…United States…world
- current birth cohorts

Expand phenotype data
- all Utah hospitals, surrounding states, U.S.-wide (VA)
- labs, images, medications, treatment response, prognosis,…

Expand to other data types

geographical locations, environmental exposures, …

Veteran Genealogy Project

Goal: Build the US genealogy, including immigrant founders and link to the 11 million veterans served by the VHA.

3 year grant completed

  Feasibility study for 2 states: Utah and Massachusetts

  38 million individuals

  linked to 500,000 veterans using VA health facilities

Grant renewed to do 2 more states

ACKNOWLEDGMENTS

Division of Genetic Epidemiology
Department of Internal Medicine
University of Utah School of Medicine

Craig Teerlink, Phd

Jim Farnham

Steve Backus